ORIGINS AND IMPLICATIONS OF THE D STAGGER IN COLLAGEN

Barbara B. Doyle, David W.L. Hukins, David J.S. Hulmes,
Andrew Miller, Christopher J. Rattew[1] and John Woodhead-Galloway

Laboratory of Molecular Biophysics
Department of Zoology
South Parks Road, Oxford OX1 3PS

Received August 15, 1974

Summary Although the distribution of hydrophobic residues in the α1 chain collagen sequence has a D ≃ 670 Å periodicity, it is dipoles formed by 100 residues occurring in pairs of unlike charge which are responsible for the 1D stagger between molecules. Sheet models based on the Hodge-Petruska model for the axially projected collagen structure require interactions specifying 1D and 4D staggers. We found no evidence for interactions specifying a strong 4D stagger and, therefore, favour the Smith microfibril model which is specified by 1D stagger interactions alone. Two hydroxylysine residues, 234 residues apart, may form a covalent cross-link stabilising the 1D stagger. Gene duplication does not appear to be responsible for the periodicity in the sequence.

Introduction The D ≃ 670 Å periodicity of collagen fibres is caused by parallel molecules of length 4.4D having an axial translation, or stagger, of nD with respect to each other, where n is an integer (1). It has been shown (2) that the distribution of charged and large hydrophobic residues in the triple-helical region, for the available α1 chain sequence, is responsible for favourable interactions at 0D, 1D, 2D, 3D and 4D staggers, where D = 234 ± 1 residue translations.[2] We shall show that the 1D stagger is specified by charged residues occurring close together in the sequence in pairs of opposite charge. Consideration of possible covalent cross-links, subsequently formed between molecules,

---

1. Present address: Department of Biophysics, University of London King's College, 26-29 Drury Lane, London WC2B 5RL.

2. Comparison of observed and computer generated electron micrographs yielded D = 232 ± 0.5 residue translations (3). Walton (4,5) obtained D = 233 ± 1 by methods similar to those of refs. 2 and 3. X-ray diffraction suggests D = 234 ± 2 (6).

provides further evidence for a 1D stagger between adjacent molecules. The 1D stagger has important implications for the formation of a collagen microfibril (7). Surprisingly we shall show that the pseudo D period in the sequence, which determines this 1D stagger (8), is not a result of gene duplication.

Hydrophobic interactions Hydrophobic residues alone lead to favourable interactions for integral staggers of D with, of course, the maximum possible number of hydrophobic interactions occurring at 0D (2) (see fig. 1a). For anti-parallel molecules, however, the most stable interactions do not occur at integral D staggers (9).

Electrostatic interactions In general any distribution of charged residues would tend to destabilise a 0D stagger, which would maximise repulsions between like charges, but would not necessarily stabilise any D stagger. We have examined the sequence to see if any particular aspect of its distribution of charged residues would favour an integral D stagger.

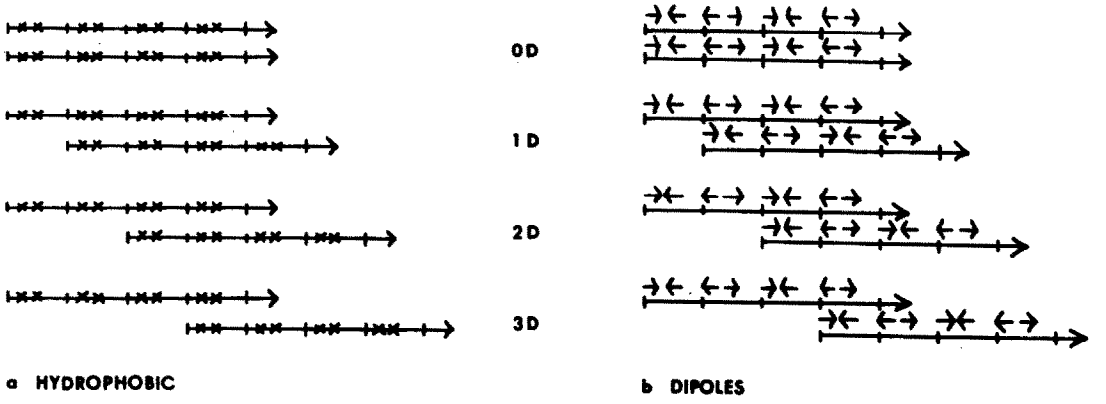

Fig. 1 Schematic diagram showing how (a) hydrophobic residues and (b) dipoles can specify D staggers. Any apparently random, but asymmetric, distribution of hydrophobic residues (crosses) which was, at least roughly, repeated in every D period would lead to a D periodicity in the interactions of parallel, but not anti-parallel, molecules. Pairing of unlike charged residues to form dipoles (arrows) can stabilise 1D and 3D staggers if the dipoles point in opposite directions in successive D periods, since parallel dipoles have a destabilising influence while anti-parallel dipoles contribute to stability. In practice, no peak corresponding to a 3D stagger appears above the noise level of fig.2. This figure does not show the actual distribution of hydrophobic residues or dipoles.

Table 1 Positions of paired charged residues classified according to the direction of the resulting dipole

1. Dipoles pointing towards C terminus

Arg-Asp	144-146	567-569	845-846	989-992	
Arg-Glu	183-186	192-194	294-296	729-731	737-740
	833-834	855-857	885-887		
Lys-Asp	50-53	264-266	434-435	531-533	564-566
	573-576	581-582	654-656	852-854	881-884
	971-972				
Lys-Glu	108-110	174-176	270-272	645-647	722-723

2. Dipoles pointing towards N terminus

Asp-Arg	384-386	620-621			
Glu-Arg	131-132	308-309	332-333	347-350	419-420
	452-453	500-501	506-507	554-555	788-789
	812-813				
Asp-Lys	96-99	372-374	602-603		
Glu-Lys	56-59	288-290	413-416	476-479	752-753
	800-803				

The triple-helical region of the α1 chain sequence (2) contains 87 positively and 73 negatively charged residues. Of these 100 occur in pairs so that a positively and a negatively charged residue are separated by ≤ 2 uncharged residues (Table 1), as has been mentioned previously (2). The remaining 60 unpaired charges are implicated in the formation of fibrous long spacing (FLS) collagen (8) but they do not specify a D periodicity.

Fig.2 shows that the most stable interaction between parallel molecules, when only the paired charge residues are considered, occurs for molecules staggered by 1D. We may consider the paired charged residues to form dipoles and fig.1b shows that this idea aids our understanding of how paired charges specify a 1D stagger.

Pairs of residues of opposite charge were previously unsuspected of having any special importance for the interactions of collagen, or any other protein, molecules and may prove to be of more general importance.

Cross-links After the triple helical collagen molecules have assembled into a fibril, under the influence of hydrophobic and electrostatic interactions, the fibril is stabilised by the formation of covalent cross-links involving lysine and hydroxylysine (10). Two of the four hydroxylysine residues in the sequence (at positions 681 and 915) are 234 residues apart, suggesting that their function is to form cross-links which stabilise the 1D stagger between adjacent molecules. Furthermore 4 of the 32 lysine residues occur in pairs 234 residues apart, while 10 occur in pairs which are 234 ± 3 residues apart.

Implications for Smith's microfibril If adjacent molecules in the most general axially projected collagen fibril model, with parallel molecules related by staggers of nD (1), are related by a 1D stagger

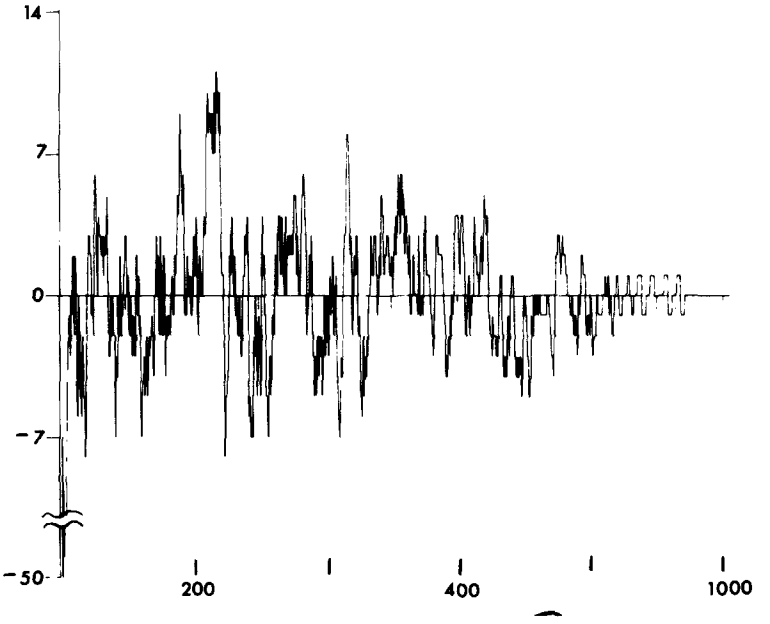

Fig.2 Interactions between dipoles on two parallel molecules as a function of the stagger between them. Dipoles consist of unlike charged residues separated by less than three residues. Anti-parallel dipoles within a range of ± 3 residues, score +1; parallel dipoles score -1. The ordinate is the total score and the abscissa is measured in residues. The 1D peak has a height corresponding to 11 stabilising dipole-dipole interactions: since the minimum number of dipoles pointing in either direction is 22 this represents 50% of the maximum possible stability.

then the familiar Hodge-Petruska model is produced (11, 12). We have found the source, in the α1 chain sequence, of interactions leading specifically to a 1D stagger and have identified residues which might form a cross-link stabilising this stagger. Therefore we prefer the Hodge-Petruska model to the other schemes, in which adjacent molecules would be related by other specific nD staggers. (see ref. 6).

The Hodge-Petruska, or indeed any other, model for the axially projected structure of the collagen fibril does not directly lead to a three-dimensional model. Staggered molecules could form sheets but Smith (7) suggested that a sheet with a width of five molecules was rolled into a cylinder (fig. 3). Evidence for (but not proof of) this cylinder, or microfibril, was provided by X-ray diffraction (13) and electron microscopy (6, 14, 15). Further evidence is provided by the results described here.

A simple sheet model like that of fig.3 requires two kinds of interactions for its integrity: one kind specifies a 1D stagger while the other specifies a 4D stagger. (Other sheet models also require at least two kinds of interactions). Although we have found the source of a strong 1D stagger we have found no such evidence for a comparably strong 4D stagger. However if this sheet is rolled into a cylinder the molecules become sub-units on a five-fold helix which can be specified solely by one interaction between successive

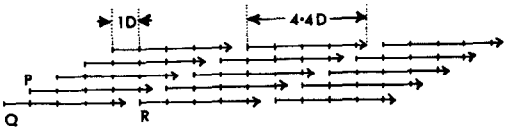

Fig.3 Hodge-Petruska model for the axially projected collagen structure. A sheet model based on this structure requires two kinds of interactions between molecules: one kind (between molecules like P and Q) leads to a 1D stagger, while the other (between molecules like P and R) leads to a 4D stagger.

residues: the 1D stagger. Once the microfibril has assembled the resulting 4D stagger may be stabilised by specific cross-links but only the interactions responsible for the 1D stagger are necessary for self-assembly.

Genetic implications We have investigated how a collagen molecule with the D periodic interactions, which we have observed, could have evolved. At first sight the obvious mechanism would be by successive duplication of a gene corresponding to a length D of sequence. We investigated this possibility by translating the sequence past itself, one residue at a time, counting the number of homologies; thus we did not allow for the possibility of deletions since we were searching for the source of an exact periodicity produced by duplication of an entire D segment with glycine as every third residue (see ref. 16). At no position, other of course than OD, did the number of identical residues exceed 15% (if gene duplication were responsible we might expect about 50% homology - ref. 17). Furthermore when amino acids related by a single base change in their codons were considered to be homologous, this value did not exceed 30%. We conclude that there is no evidence in the sequence for the D periodicity being caused primarily by gene duplication.[3]

Upon further consideration we can see that gene duplication is unlikely to be responsible for a 1D periodicity because the 1D stagger is specified by interactions between unlike charged residues. Gene duplication, although ideal for specifying the distribution of hydrophobic residues, would not produce oppositely charged residues at corresponding positions in consecutive D periods. Once the observed periodicity was established in a 2D length it is conceivable that the periodicity of the entire sequence could have been achieved

---

3. Similar conclusions were reached by C.J. Maclean and K.A.Piez (2).

by gene duplication. However, we found no evidence for this possibility and can only conclude that whatever mechanism is responsible for the 1D periodicity, within a 2D period, is probably responsible for the periodicity of the entire sequence.

In conclusion we are beginning to understand how the collagen sequence is responsible for the self-assembly of molecules into fibrils but not how this periodic sequence evolved.

Acknowledgements This work is part of the programme of the Medical Research Council Research Group in Molecular Biophysics, University of Oxford. B.B.D. is a Helen May Whitney Fellow and J. W.-G. is supported by the Arthritis and Rheumatism Council.

References

1. Cox, R.W., Grant, R.A. and Horne, R.W. (1967) J.Roy.Microscop. Soc. 87, 123-142.
2. Hulmes, D.J.S., Miller, A., Parry, D.A.D., Piez, K.A. and Woodhead-Galloway, J. (1973) J.Mol.Biol. 79, 137-148.
3. Chapman, J.A. and Hardcastle, R.A. (1974) Connective Tissue Res. in press.
4. Walton, A.G. (1973) in "Preprints of the Vth Annual Biomaterials Symposium", Hulbert, S.F. (ed.), Section 7, p.1, Clemson University, South Carolina.
5. Walton, A.G. (1974) J.Biomed.Matls.Res. in press.
6. Doyle, B.B., Hulmes, D.J.S., Miller, A., Parry D.A.D., Piez, K.A. and Woodhead-Galloway, J. (1974) Proc.Roy.Soc.Lond. Ser.B., 186, 67-74.
7. Smith, J.W. (1968) Nature, 219, 157-158.
8. Doyle, B.B., Hulmes, D.J.S., Miller, A., Parry, D.A.D., Piez, K.A. and Woodhead-Galloway, J. (1974) Proc.Roy.Soc.Lond.Ser.B., in press.
9. Doyle, B.B., Hukins, D.W.L., Hulmes, D.J.S., Miller, A. and Woodhead-Galloway, J. (1974) J.Mol.Biol. submitted for publication.
10. Tanzer, M.L. (1973) Science, 180, 561-566.
11. Schmitt, F.O., Gross, J. and Highberger, J.H. (1955) Exp. Cell Res.Suppl. 3, 326-334.
12. Hodge, A.J. and Petruska, J.A. (1963) in "Aspects of Protein Structure" Ramachandran, G.N. (ed.), pp.289-300, Academic Press, New York.
13. Miller, A. and Wray, J.S. (1971) Nature, 230, 437-439.
14. Smith, J.W. and Frame, J. (1969) J.Cell Sci. 4, 421-436.
15. Bouteille, M. and Pease, D.C. (1971) J.Ultrastructure Res. 35 314-338.
16. Needleman, S.B. and Wunsch, C.D. (1970) J.Mol.Biol. 48, 443-453.
17. Yudkin, M.and Offord, R. (1973) "Comprehensible Biochemistry", pp. 467-470, Longman, London.